Digital Imaging of Wounds: Are Measurements Reproducible Among Observers?

Susan Y. Quan, BS, Gerald S. Lazarus, MD,* Angela R. Kohli, BSN, WOCN,**
Roger Kapoor, BS,† and David J. Margolis, MD, PhD†
*Johns Hopkins Wound Center and Department of Dermatology, Johns Hopkins Bayview Medical Center, Baltimore, Maryland, and †Department of Dermatology and Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania

*Advances in digital imaging and archiving have made the measurement and documentation of wound areas possible over time. To assess the reproducibility and precision of digital image measurements, we used Wound*Matrix *Web (http://www.woundmatrix.com/) and recruited a group of caregivers from the Johns Hopkins Wound Center to measure the size of wounds on digital images by measuring length and width and tracing the circumference of the same wounds. One set of images was provided by Wound*Matrix *(Wound*Matrix *Inc, Chadds Ford, PA) and a second set used our own photographs taken at the Johns Hopkins Wound Center. Our results demonstrate that digital analysis with Wound*Matrix *Web is reproducible and precise with acceptable variation among readers. This supports the use of digital images of wounds to follow clinical progress as well as analyze the effects of new clinical interventions in clinical trials.*

Key words: Wound*Matrix*, wound healing, wound area

The use of digital imaging over the past decade has increased dramatically as advances in capture, analysis, and archiving have matured. Digital archiving permits immediate access to images so change over time can be quantified more conveniently and at lower cost. More importantly, the use of digital images, if a reproducible tool, allows clinicians to follow wound healing over time so one can determine whether specific therapeutic interventions are beneficial.

Kantor and Margolis[1] have suggested that careful observation over short periods of time (eg, 4 weeks) can predict wound healing by obtaining a first derivative of change, which mathematically demonstrates the rate of change. In a multicenter study, these observers showed that percentage of change in area over time, but not rate of healing or area healed per week, distinguished between those with venous leg ulcers that healed and those who failed to heal after 24 weeks of good wound care. That presumes that the measurements of wound size are precise and reproducible. This study analyzes how consistent individual observers are in making repeated measurements of wound area and evaluates the variability between observers in their measurements of digital images. No conclusions can be drawn, no matter how sophisticated the technology, if the observers are inconsistent in their own measurements or they vary from fellow observers.

There are several convenient methods to measure wound area in images, including (1) careful tracing of the perimeter of the wound and (2) measuring greatest length × greatest perpendicular width. In another publication, Kantor and Margolis[2] found a tight correlation between planimetric determination of wound area and simple measurements of wound length × greatest width perpendicular to length. However, the correlation between simple length versus width and planimetric tracings dropped considerably for wounds greater in area than 40 cm^2. The authors indicated that the issue was unresolved for larger wounds and irregular shaped wounds. Intuitively, across a wide diversity of irregularly shaped wounds of varying size, it would seem more reproducible to trace the

Correspondence should be sent to: Susan Y. Quan, Johns Hopkins Wound Center and Department of Dermatology, Johns Hopkins Bayview Medical Center, 4940 Eastern Avenue, Baltimore, MD 21224; e-mail: susie@jhmi.edu.

DOI: 10.1177/1534734607306880

circumference of a wound rather than to use a formula using length and width.

This study reports how precise and reproducible single observers are in measuring the same wound on multiple occasions and the variation among experienced wound care doctors and nurses in measuring the same wounds. Wound*Matrix* Web (http://www.woundmatrix.com/), a software program designed for digital measurement and documentation, was used to evaluate these issues in images supplied by Wound *Matrix* (Wound*Matrix* Inc, Chadds Ford, PA) and in images taken at the Johns Hopkins Wound Center. The data were analyzed by the clinical epidemiology unit of the Department of Dermatology at the University of Pennsylvania.

METHODS

Readers were recruited from the Johns Hopkins Wound Center and consisted of nurses, physicians, and medical students. Nine readers completed tracing the wound perimeter and measuring length and width using the image analysis tools supplied in Wound*Matrix* Web for a set of 25 distinct images. These images were provided at our request and selected by Wound*Matrix* from their in-house database of pictures acquired from several wound care provider populations to cover a wide range of wound types, sizes, shapes, and degrees of lighting. Six readers completed tracing the wound perimeter and length × width measurements for a set of 15 images, which were taken at the Johns Hopkins Wound Center of two wounds from one patient over a 1-month period of time.

All images, either provided by Wound*Matrix* Web or taken at our center, included a ruler in centimeters photographed adjacent to the wound for calibration purposes. For each image, each reader set the calibration by drawing a line parallel to the edge of the ruler of at least 4 cm in length. The reader then outlined the wound perimeter using the pen tool. The Wound*Matrix* Web program calculated the area within the traced region corresponding to the calibration value set. Without changing the calibration, each reader then selected the length button and drew a line that was the longest possible of the wound. Each reader then selected the width button and drew the longest possible line of the wound that was perpendicular to the length. Measurements were recorded of the tracings and length × width for both sets of images and submitted for statistical analysis.

Data Analysis

First, an inter-item correlation matrix was created to evaluate differences between readers. Second, 1-way analysis of variance (ANOVA) was calculated to determine if differences existed within the dataset between readers. Finally, estimates of reliability are presented as intraclass correlation coefficient (ICC). ICC was estimated using a 2-way mixed-effects model for single measures and are reported with 95% confidence intervals in parentheses. These models are based on 1-way ANOVA measures per Shrout and Fleiss.[3] Computations of reliability and all other statistical analyses were conducted using Statistical Package for the Social Sciences (SPSS), version 13 for Windows (SPSS, Inc, Chicago, IL). Constraints of the ANOVA analysis program meant that if any data were missing for a reader, then all of data respective to that reader were excluded from analysis. As several readers in the “25” group did not provide one or two estimates, many of their other observations were not part of our initial ICC estimate.

RESULTS

The table (Table 1) shows the intraclass correlation coefficients for length times width and tracing measurements for each set of images. To determine the amount of within-subject variability, we computed ICC. For the set of 25 images, the ICC for the readers’ traced images was 0.64 (0.49, 0.79). The ICC for these estimated by length × width was 0.70 (0.42, 0.94). These numbers varied little when the dataset was modified to include the greatest number of measurements. The highest ICCs were observed for the set of 15 images. The ICC for the 15 images that were traced was 0.91 (0.83, 0.96) and the ICC for the 15 images estimated by length and width measurements was 0.80 (0.65, 0.91). There were no missing data in this dataset. All P values for the ICC measurements as estimated for both sets of images were less than .0001.

Our ANOVA analysis revealed no significant statistical differences in the set of 25 images when images were traced ($P = .857$) or interpreted by length and width ($P = .453$). Results for the set of 15 images followed a similar pattern when traced ($P = .401$) and when measured by length and width ($P = .453$).

Our estimates of inter-item correlation between readers was generally high with measurements consistently greater than 0.80. However, one of nine readers of the set of 25 images consistently had correlations

Table 1. Intraclass Correlation Coefficient with 95% Confidence Interval in Parentheses

	Intraclass Correlation Coefficient With 95% Confidence Interval	Intraclass Correlation Coefficient With 95% Confidence Interval Without Reader #4
Set of 25, L × W	0.70 (0.42, 0.94)	0.72 (0.58, 0.85)
Set of 25, Tracing	0.64 (0.49, 0.79)	0.81 (0.58, 0.96)
Set of 15, L × W	0.80 (0.65, 0.91)	0.82 (0.68, 0.93)
Set of 15, Tracing	0.91 (0.83, 0.96)	0.91 (0.86, 0.99)

L = Length; W = width.

of less than 0.40 and even had scores less than 0 as compared with the other readers. One of the six readers of the set of 15 images did have correlations as low as 0.70 as compared with the other readers. When the outlier was corrected for, our correlation coefficients improved. Tracing measurements in the set of 25 images were most significantly affected (0.81 vs 0.64).

DISCUSSION

The use of digital images in a wound care setting is a major advance. The ability to refer to previous photographs and assess progress is important in defining therapeutic strategies. Patients and caregivers no longer need to make subjective judgments that compare the current status of the wound with that from a previous appointment with descriptors such as looks "better," "worse," "drier," "wetter," "pinker," and so forth. Digital analysis can be used as an analytical tool to follow wounds and it is critical that all personnel are trained and tested to ensure reliability and reproducibility. This, however, does not eliminate the need for assessment of wounds with sterile probes where undermining exists.

Our results from two sets of images, 25 distinct images provided by Wound*Matrix* Web or repeated examination of our own images of two wounds that did not change over a 1-month period of time, demonstrate that digital analysis with Wound*Matrix* Web is reproducible and precise. Furthermore, the variation among examiners who are trained and experienced is within acceptable limits for good practice. The interpretation of ICC is often based on the purpose of the investigator's research. However, many have accepted the criteria established by Landis and Koch for Cohen's kappa statistics, which is used as an estimate of intrarater correlation for categorical responses. In this setting 0.40 to 0.59 is considered moderate inter-rater reliability, 0.60 to 0.79 is considered substantial inter-rater reliability, and 0.80 and greater is considered outstanding.[4]

The set of 25 images were completed by our readers at a time when we had recently begun to use Wound*Matrix* Web. The correlation coefficients for tracing and length × width were comparable, although we had achieved higher correlation in simple measurements than in tracing. This could be attributed to the relative ease of drawing straight lines as opposed to tracing perimeters given inexperience with the new software. The set of 15 images was completed at a later time by readers who were much more attuned to Wound*Matrix* Web and experienced in taking measurements. This was clearly demonstrated in the increase in correlation coefficients for both tracing and length × width when comparing the set of 25 images to the set of 15 images.

Reader 4's ability to measure wounds was markedly different than all of the other readers in the set of 25, and once this reader was removed from data analysis the correlations improved. Interestingly, no one in the set of 15 appeared to correlate so poorly with the other readers. This serves to illustrate that even with multiple people making measurements, one person may not be able to properly perform tracing measurements, which could put the reproducibility of the results of a study in jeopardy. When conducting a clinical trial as well as in routine office practice it is worthwhile to train all observers and confirm their ability to trace or perform simple measurements to maximize reliability of measurement. It may even be important to periodically retest competence in measuring wounds.

The use of digital image analysis is not to have an ideal measure of change in wound area but rather to have a reliable measure using technology that is now commonplace. Although our data do not tell us the true size of the wound, it shows us that we can reliably quantify change over time. Percentage change in area over time is a significant prognostic measure of complete wound healing.[2] Margolis et al showed that by creating complex models using logistic regression, as well as by simply counting prognostic factors, initial measures of wound size and duration accurately predict who will heal by the 24th week of care.[5] For example, they demonstrated that a wound that is less than 10 cm^2 in size and less than 12 months old at the first visit has a 29% chance of not healing by the 24th week of care, whereas a wound greater than 10 cm^2 in size and greater than 12 months old has a 78% chance of not healing.

The use of such models and the reproducibility of wound area measurements using Wound*Matrix* Web can be applied by a clinician to help determine when therapy is effective or not so that alternative interventions are in order. Digital image analysis using predictive rules may also have great utility in clinical trials.

ACKNOWLEDGMENTS

We thank everyone who participated in wound measurements: Penny Bragg, Gail Cockrell, Barbara Cook, Robin Dapp, Yelena Frankel, Rose Housely, Phyllis Keys, Monica Sanchez, Evelyn Taylor, and Fran Valle. We also acknowledge the support of Wound*Matrix* in providing the software at reduced cost, cooperation in programming the system to our specifications, and providing test images. No honoraria or additional support was provided.

REFERENCES